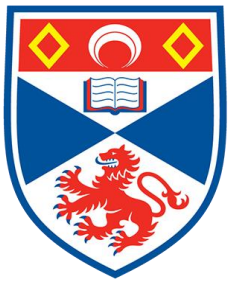
Statistics, Machine Learning and Deep Learning for Population Genetic Inference

Xinghu Qin



University of
St Andrews

This thesis is submitted in partial fulfilment for the degree of

Doctor of Philosophy (PhD)

at the University of St Andrews

February 2021

Candidate's declaration

I, Xinghu Qin, do hereby certify that this thesis, submitted for the degree of PhD, which is approximately 68, 352 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree. I confirm that any appendices included in my thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

I was admitted as a research student at the University of St Andrews in August 2016.

I received funding from an organisation or institution and have acknowledged the funder(s) in the full text of my thesis.

Date 27/05/2021 Signature of candidate

Supervisor's declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree. I confirm that any appendices included in the thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

Date May 27, 2021 Signature of supervisor

Permission for publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Xinghu Qin, have obtained, or am in the process of obtaining, third-party copyright permissions that are required or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

Printed copy

Embargo on part (Chapter 1, Chapter 2, Chapter 3, Chapter 4, Chapter 5, Appendix A, B, and C.) of print copy for a period of 1 years on the following ground(s):

- Publication would preclude future publication

Supporting statement for printed embargo request

I have a publishing contract pending

Electronic copy

Embargo on part (Chapter 1, Chapter 2, Chapter 3, Chapter 4, Chapter 5, Appendix A, B, and C.) of electronic copy for a period of 1 years on the following ground(s):

- Publication would preclude future publication

Supporting statement for electronic embargo request

I have a publishing contract pending

Title and Abstract

- I agree to the title and abstract being published.

Date 27/05/2021 Signature of candidate

Date May 27, 2021 Signature of supervisor

Underpinning Research Data or Digital Outputs

Candidate's declaration

I, Xinghu Qin, hereby certify that no requirements to deposit original research data or digital outputs apply to this thesis and that, where appropriate, secondary data used have been referenced in the full text of my thesis.

Date 27/05/2021 Signature of candidate

Acknowledgements

I would like, first and foremost, to express my deepest and most sincere thanks to my supervisor Prof Oscar E Gaggiotti. Thank you for your guidance and patience to make me grow over the last four years. Thank you for your academic advice and supervision to complete my PhD thesis.

The members of the MEME group (Gaggiotti Lab) have shared, provided generosity and patience in terms of science and life over the last four years: Dr Valeria Montano, Dr Marie Louis, Mrs Aubrie Booth, Mrs Eadin O'mahony. Thanks for your comfort and encouragement or otherwise put up with me.

Thanks Dr Charleston wen-kai Chiang for his help to get access to the POPRES data and USC cluster, as well as for his comments on the third chapter.

Moreover, I would like to express my thanks to whom I have learnt important scientific knowledge and techniques, as well as life experiences. There are so many people that I cannot name everyone in the list (chronological). Jane Brooks at ELT (IEI) has provided patient, considerate, and persistent help from academic English learning to life advice. Mr Sturt Newland and Ms. Jane Newland, who provided place to share delicious food, the knowledge of God, life, and also travel opportunities to see Scotland during the past four years. The fellowship, encouragement, and bath of the soul from (and with) Fiona and Ian, Chuyin, Wenbo, Linda, ... The accompany and unforgettable memory from Mary.

Thanks for staff and students in SOI, CBD and whole school of biology, thanks the University of St Andrews for providing me the opportunity and place to study and work in Scotland.

Last but not least, my greatest thanks to my parents. My degree is not their interests, but they always love me greatly and wish me happy and safe.

Funding support

This work was supported by China Scholarship Council -University of St Andrews Scholarship.

Ethical Statement

This research did not have direct animal involvement or human participant. The access, storage and usage of the human genetic data (POPRES and CONVERGE) were approved by the School of Biology Ethics Committee. The research was conducted following the regulations of University Teaching and Research Ethics Committee (UTREC). The POPRES data was applied and accessed under the support of Dr Charleston Chiang from University of Southern California with the application approved by the USC Human Research Ethics Committee (HREC).

Contents

Abstract

Deciphering the evolutionary changes from raw DNA data effectively without the loss of intrinsic information has been the fundamental and core work in population genetics. However, some statistical challenges still restrict the inferential performance in population genetics, for example, the undue emphasis on rare or common alleles measured by different statistics, the ubiquitous multimodal genetic structure within populations, and complex genotype-by-environment associations. In this thesis, I propose to integrate the information-based statistics with machine learning approaches to address these problems and challenges for population genetic inference. First, I evaluated the performance of the information-based summary statistics for spatial demography inference. I showed that the summary statistics based on Shannon differentiation and the transformed diversity of order q=1 had higher power to discriminate spatially-structured scenarios than the traditional allelic richness and heterozygosity-based summary statistics. This provides guidelines for using summary statistics to make inference of spatial demography and for developing new statistical methods to detect signatures of evolutionary changes. Second, I proposed to use Kernel Local Fisher Discriminant Analysis of Principal Components (KLFDAPC) for population genetic structure inference considering the nonlinear and multimodal genetic information between individuals. KLFDAPC outperformed both PCA and DAPC in discriminatory power and in predicting individual geographic origin. KLFDAPC is useful for geographic ancestry inference and correction of population stratification in GWAS. Finally, I proposed a deep learning-based approach (*DeepGenomeScan*) to detect signals of selection. *DeepGenomeScan* had higher power than the commonly used machine learning approaches such as *pcadapt* and *RDA* in identifying signatures of selection. Furthermore, *DeepGenomeScan* can be extended to implement various genome-wide association studies (GWAS, TWAS, PWAS, and MWAS) by performing a systematic scanning on genome-wide variants to detect the genetic variations responsible for complex traits or involved in adaptation. In summary, this dissertation addresses several foundational questions in statistics-based and machine learning-based inference, contributing several the-state-of-the-art statistical tools for population genetic inference.

Chapters 1-5 and
Appendices A, B, and C
have been embargoed at
the author's request.

References

ABADI, M., BARHAM, P., CHEN, J., CHEN, Z., DAVIS, A., DEAN, J., DEVIN, M., GHEMAWAT, S., IRVING, G. & ISARD, M. Tensorflow: A system for large-scale machine learning. 12th USENIX symposium on operating systems design and implementation (OSDI 16), 2016. 265-283.

ABDELLAOUI, A., HUGH-JONES, D., YENGO, L., KEMPER, K. E., NIVARD, M. G., VEUL, L., HOLTZ, Y., ZIETSCH, B. P., FRAYLING, T. M. & WRAY, N. R. 2019. Genetic correlates of social stratification in Great Britain. *Nature human behaviour,* 3**,** 1332-1342.

ABDI, H. & WILLIAMS, L. J. 2010. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics,* 2**,** 433-459.

AHMED, O. & BRIFCANI, A. Gene Expression Classification Based on Deep Learning. 2019 4th Scientific International Conference Najaf (SICN), 2019. IEEE, 145-149.

AITKEN, S. N., LUIKART, G. & ALLENDORF, F. W. 2012. *Conservation and the genetics of populations*, John Wiley & Sons.

ÅKESSON, M., SINGH, P., WREDE, F. & HELLANDER, A. 2020. Convolutional Neural Networks as Summary Statistics for Approximate Bayesian Computation. *arXiv preprint arXiv:2001.11760*.

AL BKHETAN, Z., KADLOF, M., KRAFT, A. & PLEWCZYNSKI, D. 2019. Machine Learning Polymer Models of Three-Dimensional Chromatin Organization in Human Lymphoblastoid Cells.

ALANIS-LOBATO, G., CANNISTRACI, C. V., ERIKSSON, A., MANICA, A. & RAVASI, T. 2015. Highlighting nonlinear patterns in population genetics datasets. *Scientific reports,* 5**,** 8140.

ALEXANDER, D. H., NOVEMBRE, J. & LANGE, K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome research,* 19**,** 1655-1664.

ALIPANAHI, B., DELONG, A., WEIRAUCH, M. T. & FREY, B. J. 2015. Predicting the sequence specificities of DNA-and RNA-binding proteins by deep learning. *Nature biotechnology,* 33**,** 831-838.

ALLEN, B., KON, M. & BAR-YAM, Y. 2009. A new phylogenetic diversity measure generalizing the Shannon index and its application to phyllostomid bats. *The American Naturalist,* 174**,** 236-243.

ALLENDORF, F. W. 1986. Genetic drift and the loss of alleles versus heterozygosity. *Zoo biology,* 5**,** 181-190.

ALONSO-BLANCO, C., ANDRADE, J., BECKER, C., BEMM, F., BERGELSON, J., BORGWARDT, K. M., CAO, J., CHAE, E., DEZWAAN, T. M. & DING, W. 2016. 1,135 genomes reveal the global pattern of polymorphism in Arabidopsis thaliana. *Cell,* 166**,** 481-491.

ALVARADO‐SERRANO, D. F. & HICKERSON, M. J. 2016. Spatially explicit summary statistics for historical population genetic inference. *Methods in Ecology and Evolution,* 7**,** 418-427.

AMIR, S., WALLACE, B. C., LYU, H. & SILVA, P. C. M. J. 2016. Modelling context with user embeddings for sarcasm detection in social media. *arXiv preprint arXiv:1607.00976*.

ANGERMUELLER, C., LEE, H. J., REIK, W. & STEGLE, O. 2017. DeepCpG: accurate prediction of single-cell DNA methylation states using deep learning. *Genome biology,* 18**,** 1-13.

ANGERMUELLER, C., PÄRNAMAA, T., PARTS, L. & STEGLE, O. 2016. Deep learning for computational biology. *Molecular systems biology,* 12.

BABAUD, J., WITKIN, A. P., BAUDIN, M. & DUDA, R. O. 1986. Uniqueness of the Gaussian kernel for scale-space filtering. *IEEE transactions on pattern analysis and machine intelligence***,** 26-33.

BAKER, M. R. & PATIL, R. B. 1998. Universal approximation theorem for interval neural networks. *Reliable Computing,* 4**,** 235-239.

BALDING, D. J. 2003. Likelihood-based inference for genetic correlation coefficients. *Theoretical population biology,* 63**,** 221-230.

BARBUJANI, G. & EXCOFFIER, L. G. L. 1999. *The history and geography of human genetic diversity*, Oxford University Press.

BATTEY, C., COFFING, G. C. & KERN, A. D. 2020a. Visualizing population structure with variational autoencoders. *BioRxiv*.

BATTEY, C. J., RALPH, P. L. & KERN, A. D. 2020b. Predicting geographic location from genetic variation with deep neural networks. *ELife,* 9**,** e54507.
BAUM, E. B. 1988. On the capabilities of multilayer perceptrons. *Journal of complexity,* 4**,** 193-215.
BAZIN, E., DAWSON, K. J. & BEAUMONT, M. A. 2010. Likelihood-free inference of population structure and local adaptation in a Bayesian hierarchical model. *Genetics,* 185**,** 587-602.
BEALE, C. M., LENNON, J. J., YEARSLEY, J. M., BREWER, M. J. & ELSTON, D. A. 2010. Regression analysis of spatial data. *Ecology letters,* 13**,** 246-264.
BEAUMONT, M. A., ZHANG, W. & BALDING, D. J. 2002. Approximate Bayesian computation in population genetics. *Genetics,* 162**,** 2025-2035.
BELKINA, A. C., CICCOLELLA, C. O., ANNO, R., HALPERT, R., SPIDLEN, J. & SNYDER-CAPPIONE, J. E. 2019. Automated optimized parameters for T-distributed stochastic neighbor embedding improve visualization and analysis of large datasets. *Nature communications,* 10**,** 1-12.
BELLOT, P., DE LOS CAMPOS, G. & PÉREZ-ENCISO, M. 2018. Can deep learning improve genomic prediction of complex human traits? *Genetics,* 210**,** 809-819.
BENESTAN, L., GOSSELIN, T., PERRIER, C., SAINTE-MARIE, B., ROCHETTE, R. & BERNATCHEZ, L. 2015. RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the A merican lobster (H omarus americanus). *Molecular ecology,* 24**,** 3299-3315.
BERGMANN, F., GREGORIUS, H.-R., KOWNATZKI, D. & WEHENKEL, C. 2013. Different diversity measures and genetic traits reveal different species genetic diversity relationships: A case study in forest tree communities. *Silvae Genetica,* 62**,** 25-37.
BERGMEIR, C. N. & BENÍTEZ SÁNCHEZ, J. M. Neural networks in R using the Stuttgart neural network simulator: RSNNS. 2012. American Statistical Association.
BHAGWAT, M. 2010. Searching NCBI's dbSNP database. *Current protocols in bioinformatics,* 32**,** 1.19. 1-1.19. 18.
BHATIA, G., PATTERSON, N., SANKARARAMAN, S. & PRICE, A. L. 2013. Estimating and interpreting FST: the impact of rare variants. *Genome research,* 23**,** 1514-1521.
BLUM, M. G. & FRANÇOIS, O. 2010. Non-linear regression models for Approximate Bayesian Computation. *Statistics and computing,* 20**,** 63-73.
BLUM, M. G., NUNES, M. A., PRANGLE, D. & SISSON, S. A. 2013. A comparative review of dimension reduction methods in approximate Bayesian computation. *Statistical Science,* 28**,** 189-208.
BONHOMME, M., CHEVALET, C., SERVIN, B., BOITARD, S., ABDALLAH, J., BLOTT, S. & SANCRISTOBAL, M. 2010. Detecting selection in population trees: the Lewontin and Krakauer test extended. *Genetics,* 186**,** 241-262.
BOTTA, V., LOUPPE, G., GEURTS, P. & WEHENKEL, L. 2014. Exploiting SNP correlations within random forest for genome-wide association studies. *PloS one,* 9**,** e93379.
BRADLEY, R. A. & TERRY, M. E. 1952. Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika,* 39**,** 324-345.
BREIMAN, L. 2001. Random forests. *Machine learning,* 45**,** 5-32.
BREIMAN, L., FRIEDMAN, J., OLSHEN, R. & STONE, C. 1984. Classification and Regression Trees.
BRIEUC, M. S., ONO, K., DRINAN, D. P. & NAISH, K. A. 2015. Integration of Random Forest with population-based outlier analyses provides insight on the genomic basis and evolution of run timing in Chinook salmon (Oncorhynchus tshawytscha). *Molecular Ecology,* 24**,** 2729-2746.
BRINKER, K. Incorporating diversity in active learning with support vector machines. Proceedings of the 20th international conference on machine learning (ICML-03), 2003. 59-66.
BRO, R. & SMILDE, A. K. 2014. Principal component analysis. *Analytical Methods,* 6**,** 2812-2831.
BRYNEDAL, B., WOJCIK, J., ESPOSITO, F., DEBAILLEUL, V., YAOUANQ, J., MARTINELLI-BONESCHI, F., EDAN, G., COMI, G., HILLERT, J. & ABDERRAHIM, H. 2010. MGAT5 alters the severity of multiple sclerosis. *Journal of neuroimmunology,* 220**,** 120-124.

BUETTNER, F., PRATANWANICH, N., MARIONI, J. C. & STEGLE, O. 2016. Scalable latent-factor models applied to single-cell RNA-seq data separate biological drivers from confounding effects. *BioRxiv*, 087775.
BÜYÜKÖZTÜRK, Ş. & ÇOKLUK-BÖKEOĞLU, Ö. 2008. Discriminant Function Analysis: Concept and Application. *Eurasian Journal of Educational Research (EJER)*.
BYLANDER, T. 2002. Estimating generalization error on two-class datasets using out-of-bag estimates. *Machine learning,* 48**,** 287-297.
BYUN, J., HAN, Y., GORLOV, I. P., BUSAM, J. A., SELDIN, M. F. & AMOS, C. I. 2017. Ancestry inference using principal component analysis and spatial analysis: a distance-based analysis to account for population substructure. *BMC genomics,* 18**,** 789.
CABIN, R. J. & MITCHELL, R. J. 2000. To Bonferroni or not to Bonferroni: when and how are the questions. *Bulletin of the Ecological Society of America,* 81**,** 246-248.
CAI, N., BIGDELI, T. B., KRETZSCHMAR, W., LI, Y., LIANG, J., SONG, L., HU, J., LI, Q., JIN, W. & HU, Z. 2015. Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature,* 523**,** 588-591.
CAI, N., BIGDELI, T. B., KRETZSCHMAR, W. W., LI, Y., LIANG, J., HU, J., PETERSON, R. E., BACANU, S., WEBB, B. T. & RILEY, B. 2017. 11,670 whole-genome sequences representative of the Han Chinese population from the CONVERGE project. *Scientific data,* 4**,** 1-14.
CANDEL, A., PARMAR, V., LEDELL, E. & ARORA, A. 2016. Deep learning with H2O. *H2O. ai Inc*.
CAPBLANCQ, T., LUU, K., BLUM, M. G. & BAZIN, E. 2018. Evaluation of redundancy analysis to identify signatures of local adaptation. *Molecular ecology resources,* 18**,** 1223-1233.
CARL, G. & KÜHN, I. 2007. Analyzing spatial autocorrelation in species distributions using Gaussian and logit models. *ecological modelling,* 207**,** 159-170.
CARUANA, R. & NICULESCU-MIZIL, A. An empirical comparison of supervised learning algorithms. Proceedings of the 23rd international conference on Machine learning, 2006. 161-168.
CHANG, P., GRINBAND, J., WEINBERG, B., BARDIS, M., KHY, M., CADENA, G., SU, M.-Y., CHA, S., FILIPPI, C. & BOTA, D. 2018. Deep-learning convolutional neural networks accurately classify genetic mutations in gliomas. *American Journal of Neuroradiology,* 39**,** 1201-1207.
CHAO, A., CHIU, C.-H. & JOST, L. 2014. Unifying species diversity, phylogenetic diversity, functional diversity, and related similarity and differentiation measures through Hill numbers. *Annual Review of Ecology, Evolution, and Systematics,* 45**,** 297-324.
CHAO, A. & CHIU, C. H. 2016. Bridging the variance and diversity decomposition approaches to beta diversity via similarity and differentiation measures. *Methods in Ecology and Evolution,* 7**,** 919-928.
CHARLESWORTH, D. 2006. Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genet,* 2**,** e64.
CHEN, Y., LI, Y., NARAYAN, R., SUBRAMANIAN, A. & XIE, X. 2016. Gene expression inference with deep learning. *Bioinformatics,* 32**,** 1832-1839.
CHENG, J., LIU, H., WANG, F., LI, H. & ZHU, C. 2015. Silhouette analysis for human action recognition based on supervised temporal t-SNE and incremental learning. *Ieee transactions on image processing,* 24**,** 3203-3217.
CHIANG, C. W., MANGUL, S., ROBLES, C. & SANKARARAMAN, S. 2018. A comprehensive map of genetic variation in the world's largest ethnic group—Han Chinese. *Molecular biology and evolution,* 35**,** 2736-2750.
CHOLLET, F. 2015. Keras: Deep learning library for theano and tensorflow. *URL: https://keras. io/k,* 7**,** T1.
CHRISTIN, S., HERVET, E. & LECOMTE, N. 2019. Applications for deep learning in ecology. *Methods in Ecology and Evolution,* 10**,** 1632-1644.
CLARKE, B., FOKOUE, E. & ZHANG, H. H. 2009. *Principles and theory for data mining and machine learning*, Springer Science & Business Media.

CLEARY, J. G. & TRIGG, L. E. 1995. K*: An instance-based learner using an entropic distance measure. *Machine Learning Proceedings 1995.* Elsevier.

COHEN, J. 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement,* 20**,** 37-46.

COMUZZIE, A. G., COLE, S. A., LASTON, S. L., VORUGANTI, V. S., HAACK, K., GIBBS, R. A. & BUTTE, N. F. 2012. Novel genetic loci identified for the pathophysiology of childhood obesity in the Hispanic population. *PloS one,* 7**,** e51954.

COOP, G., WITONSKY, D., DI RIENZO, A. & PRITCHARD, J. K. 2010. Using environmental correlations to identify loci underlying local adaptation. *Genetics,* 185**,** 1411-1423.

CROSSA, J., MARTINI, J. W. R., GIANOLA, D., PEREZ-RODRIGUEZ, P., JARQUIN, D., JULIANA, P., MONTESINOS-LOPEZ, O. A. & CUEVAS, J. 2019. Deep Kernel and Deep Learning for Genome-Based Prediction of Single Traits in Multienvironment Breeding Trials.

CSILLÉRY, K., BLUM, M. G., GAGGIOTTI, O. E. & FRANÇOIS, O. 2010. Approximate Bayesian computation (ABC) in practice. *Trends in ecology & evolution,* 25**,** 410-418.

DAS, S., LECOURS BOUCHER, X., ROGERS, C., MAKOWSKI, C., CHOUINARD-DECORTE, F., OROS KLEIN, K., BECK, N., RIOUX, P., BROWN, S. T. & MOHADDES, Z. 2018. Integration of "omics" data and phenotypic data within a unified extensible multimodal framework. *Frontiers in Neuroinformatics,* 12**,** 91.

DAY, T. 2015. Information entropy as a measure of genetic diversity and evolvability in colonization. *Molecular ecology,* 24**,** 2073-2083.

DE'ATH, G. & FABRICIUS, K. E. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology,* 81**,** 3178-3192.

DE VILLEMEREUIL, P., FRICHOT, É., BAZIN, É., FRANÇOIS, O. & GAGGIOTTI, O. E. 2014. Genome scan methods against more complex models: when and how much should we trust them? *Molecular ecology,* 23**,** 2006-2019.

DE VILLEMEREUIL, P. & GAGGIOTTI, O. E. 2015. A new FST‐based method to uncover local adaptation using environmental variables. *Methods in Ecology and Evolution,* 6**,** 1248-1258.

DEL MORAL, P., DOUCET, A. & JASRA, A. 2012. On adaptive resampling strategies for sequential Monte Carlo methods. *Bernoulli,* 18**,** 252-278.

DENG, L. & LIU, Y. 2018. *Deep learning in natural language processing*, Springer.

DEPERI, S. I., TAGLIOTTI, M. E., BEDOGNI, M. C., MANRIQUE-CARPINTERO, N. C., COOMBS, J., ZHANG, R., DOUCHES, D. & HUARTE, M. A. 2018. Discriminant analysis of principal components and pedigree assessment of genetic diversity and population structure in a tetraploid potato panel using SNPs. *PloS one,* 13**,** e0194398.

DIAZ-PAPKOVICH, A., ANDERSON-TROCMÉ, L. & GRAVEL, S. 2019. UMAP reveals cryptic population structure and phenotype heterogeneity in large genomic cohorts. *PLoS genetics,* 15**,** e1008432.

DÍAZ-URIARTE, R. & DE ANDRES, S. A. 2006. Gene selection and classification of microarray data using random forest. *BMC bioinformatics,* 7**,** 3.

DICHGANS, M., MALIK, R., KÖNIG, I. R., ROSAND, J., CLARKE, R., GRETARSDOTTIR, S., THORLEIFSSON, G., MITCHELL, B. D., ASSIMES, T. L. & LEVI, C. 2014. Shared genetic susceptibility to ischemic stroke and coronary artery disease: a genome-wide analysis of common variants. *Stroke,* 45**,** 24-36.

DONG, Y., ZHUANG, B., ZHAO, L., SUN, H. & HE, M. 2001. The genetic diversity of annual wild soybeans grown in China. *Theoretical and Applied Genetics,* 103**,** 98-103.

DOUGLAS, S. C. & YU, J. Why RELU Units Sometimes Die: Analysis of Single-Unit Error Backpropagation in Neural Networks. 2018 52nd Asilomar Conference on Signals, Systems, and Computers, 2018. IEEE, 864-868.

DUARTE, J. L., CRAWFORD, J. T., STERN, C., HAIDT, J., JUSSIM, L. & TETLOCK, P. E. 2015. Political diversity will improve social psychological science 1. *Behavioral and Brain Sciences,* 38.

DUFORET-FREBOURG, N., LUU, K., LAVAL, G., BAZIN, E. & BLUM, M. G. 2016. Detecting genomic signatures of natural selection with principal component analysis: application to the 1000 genomes data. *Molecular biology and evolution,* 33**,** 1082-1093.

DUGAS, C., BENGIO, Y., BÉLISLE, F., NADEAU, C. & GARCIA, R. Incorporating second-order functional knowledge for better option pricing. Advances in neural information processing systems, 2001. 472-478.

EFRON, B. 1982. *The jackknife, the bootstrap and other resampling plans*, SIAM.

ELHAIK, E. & GRAUR, D. 2020. On the Inapplicability of Supervised Machine Learning to Evolutionary Studies. Preprints.org.

ERASLAN, G., AVSEC, Ž., GAGNEUR, J. & THEIS, F. J. 2019a. Deep learning: new computational modelling techniques for genomics. *Nature Reviews Genetics,* 20**,** 389-403.

ERASLAN, G., SIMON, L. M., MIRCEA, M., MUELLER, N. S. & THEIS, F. J. 2019b. Single-cell RNA-seq denoising using a deep count autoencoder. *Nature communications,* 10**,** 1-14.

ESTEVA, A., KUPREL, B., NOVOA, R. A., KO, J., SWETTER, S. M., BLAU, H. M. & THRUN, S. 2017. Dermatologist-level classification of skin cancer with deep neural networks. *Nature,* 542**,** 115.

ESTEVA, A., ROBICQUET, A., RAMSUNDAR, B., KULESHOV, V., DEPRISTO, M., CHOU, K., CUI, C., CORRADO, G., THRUN, S. & DEAN, J. 2019. A guide to deep learning in healthcare. *Nature medicine,* 25**,** 24-29.

EUGSTER, M. J., HOTHORN, T. & LEISCH, F. 2008. Exploratory and inferential analysis of benchmark experiments.

EXCOFFIER, L., DUPANLOUP, I., HUERTA-SÁNCHEZ, E., SOUSA, V. C. & FOLL, M. 2013. Robust demographic inference from genomic and SNP data. *PLoS Genet,* 9**,** e1003905.

EXCOFFIER, L. & FOLL, M. 2011. Fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics,* 27**,** 1332-1334.

EXCOFFIER, L., HOFER, T. & FOLL, M. 2009. Detecting loci under selection in a hierarchically structured population. *Heredity,* 103**,** 285-298.

FAUZAN, M. A. & MURFI, H. 2018. The accuracy of XGBoost for insurance claim prediction. *Int. J. Adv. Soft Comput. Appl,* 10.

FISHER, R. A. 1936. The use of multiple measurements in taxonomic problems. *Annals of eugenics,* 7**,** 179-188.

FLAGEL, L., BRANDVAIN, Y. & SCHRIDER, D. R. 2019. The Unreasonable Effectiveness of Convolutional Neural Networks in Population Genetic Inference. *Mol Biol Evol,* 36**,** 220-238.

FOLL, M. & GAGGIOTTI, O. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics,* 180**,** 977-993.

FORESTER, B. R., JONES, M. R., JOOST, S., LANDGUTH, E. L. & LASKY, J. R. 2016. Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes.

FORESTER, B. R., LASKY, J. R., WAGNER, H. H. & URBAN, D. L. 2018. Comparing methods for detecting multilocus adaptation with multivariate genotype-environment associations.

FORTUNA, M. A., ALBALADEJO, R. G., FERNÁNDEZ, L., APARICIO, A. & BASCOMPTE, J. 2009. Networks of spatial genetic variation across species. *Proceedings of the National Academy of Sciences,* 106**,** 19044-19049.

FOX, C. S., LIU, Y., WHITE, C. C., FEITOSA, M., SMITH, A. V., HEARD-COSTA, N., LOHMAN, K., JOHNSON, A. D., FOSTER, M. C. & GREENAWALT, D. M. 2012. Genome-wide association for abdominal subcutaneous and visceral adipose reveals a novel locus for visceral fat in women. *PLoS Genet,* 8**,** e1002695.

FRANÇOIS, O. & JAY, F. 2020. Factor analysis of ancient population genomic samples. *Nature communications,* 11**,** 1-11.

FRANTZ, A., CELLINA, S., KRIER, A., SCHLEY, L. & BURKE, T. 2009. Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: clusters or isolation by distance? *Journal of Applied Ecology,* 46**,** 493-505.

FRICHOT, E., SCHOVILLE, S. D., BOUCHARD, G. & FRANÇOIS, O. 2013. Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular biology and evolution,* 30**,** 1687-1699.

FUNK, W. C., LOVICH, R. E., HOHENLOHE, P. A., HOFMAN, C. A., MORRISON, S. A., SILLETT, T. S., GHALAMBOR, C. K., MALDONADO, J. E., RICK, T. C. & DAY, M. D. 2016. Adaptive divergence despite strong genetic drift: genomic analysis of the evolutionary mechanisms causing genetic differentiation in the island fox (Urocyon littoralis). *Molecular ecology,* 25**,** 2176-2194.

GAGGIOTTI, O. E., CHAO, A., PERES‐NETO, P., CHIU, C. H., EDWARDS, C., FORTIN, M. J., JOST, L., RICHARDS, C. M. & SELKOE, K. A. 2018. Diversity from genes to ecosystems: A unifying framework to study variation across biological metrics and scales. *Evolutionary Applications*.

GALINSKY, K. J., LOH, P.-R., MALLICK, S., PATTERSON, N. J. & PRICE, A. L. 2016. Population structure of UK Biobank and ancient Eurasians reveals adaptation at genes influencing blood pressure. *The American Journal of Human Genetics,* 99**,** 1130-1139.

GARSON, D. G. 1991. Interpreting neural network connection weights. *Artificial Intelligence Expert,* 6**,** 46-51.

GEETHA, A. & NASIRA, G. Data mining for meteorological applications: Decision trees for modeling rainfall prediction. 2014 IEEE international conference on computational intelligence and computing research, 2014. IEEE, 1-4.

GERRER, C.-E. & GIRARD, S. Non linear dimension reduction of dynamic model output. Proceedings of the 13th International Modelica Conference, Regensburg, Germany, March 4–6, 2019, 2019. Linköping University Electronic Press.

GEVREY, M., DIMOPOULOS, I. & LEK, S. 2003. Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological modelling,* 160**,** 249-264.

GHOSH, A. K. 2006. On optimum choice of k in nearest neighbor classification. *Computational Statistics & Data Analysis,* 50**,** 3113-3123.

GLIGORIJEVIĆ, V., BAROT, M. & BONNEAU, R. 2018. deepNF: deep network fusion for protein function prediction. *Bioinformatics,* 34**,** 3873-3881.

GOLDBERG, Y. 2017. Neural network methods for natural language processing. *Synthesis Lectures on Human Language Technologies,* 10**,** 1-309.

GOODALL, C. 1991. Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society: Series B (Methodological),* 53**,** 285-321.

GOODFELLOW, I. A. 2016. Deep learning / Ian Goodfellow, Yoshua Bengio and Aaron Courville. Cambridge, Massachusetts :: The MIT Press.

GOTSOPOULOS, A., SAARIMÄKI, H., GLEREAN, E., JÄÄSKELÄINEN, I. P., SAMS, M., NUMMENMAA, L. & LAMPINEN, J. 2018. Reproducibility of importance extraction methods in neural network based fMRI classification. *NeuroImage,* 181**,** 44-54.

GRANOVSKY, M., FATA, J., PAWLING, J., MULLER, W. J., KHOKHA, R. & DENNIS, J. W. 2000. Suppression of tumor growth and metastasis in Mgat5-deficient mice. *Nature medicine,* 6**,** 306-312.

GREENBAUM, G., TEMPLETON, A. R. & BAR-DAVID, S. 2016. Inference and analysis of population structure using genetic data and network theory. *Genetics,* 202**,** 1299-1312.

GREENBAUM, G., TEMPLETON, A. R., ZARMI, Y. & BAR-DAVID, S. 2014. Allelic richness following population founding events–a stochastic modeling framework incorporating gene flow and genetic drift. *PloS one,* 9**,** e115203.

GUARNIERI, S., PIAZZA, F. & UNCINI, A. 1999. Multilayer feedforward networks with adaptive spline activation function. *IEEE Transactions on Neural Networks,* 10**,** 672-683.

GÜNTHER, F. & FRITSCH, S. 2010. neuralnet: Training of neural networks. *The R journal,* 2**,** 30-38.

HAMILTON, G., CURRAT, M., RAY, N., HECKEL, G., BEAUMONT, M. & EXCOFFIER, L. 2005. Bayesian estimation of recent migration rates after a spatial expansion. *Genetics,* 170**,** 409-417.

HAMPE, J., SCHREIBER, S. & KRAWCZAK, M. 2003. Entropy-based SNP selection for genetic association studies. *Human Genetics,* 114**,** 36-43.
HAN, J. & MORAGA, C. The influence of the sigmoid function parameters on the speed of backpropagation learning. International Workshop on Artificial Neural Networks, 1995. Springer, 195-201.
HARDT, M., RECHT, B. & SINGER, Y. Train faster, generalize better: Stability of stochastic gradient descent. International Conference on Machine Learning, 2016. PMLR, 1225-1234.
HE, F., WANG, D., INNOKENTEVA, Y., KHOLOD, O., SHIN, D. & XU, D. 2019. Extracting Molecular Entities and Their Interactions from Pathway Figures Based on Deep Learning. 1191-1193.
HE, Q., BERTNESS, M. D., BRUNO, J. F., LI, B., CHEN, G., COVERDALE, T. C., ALTIERI, A. H., BAI, J., SUN, T. & PENNINGS, S. C. 2014. Economic development and coastal ecosystem change in China. *Scientific reports,* 4**,** 1-9.
HE, S. 2015. *Topological optimisation of artificial neural networks for financial asset forecasting.* The London School of Economics and Political Science (LSE).
HECHTER, M. 1978. Group formation and the cultural division of labor. *American Journal of Sociology,* 84**,** 293-318.
HEIJUNGS, R., HENRIKSSON, P. J. & GUINÉE, J. B. 2016. Measures of difference and significance in the era of computer simulations, meta-analysis, and big data. *Entropy,* 18**,** 361.
HELLENTHAL, G., BUSBY, G. B., BAND, G., WILSON, J. F., CAPELLI, C., FALUSH, D. & MYERS, S. 2014. A genetic atlas of human admixture history. *Science,* 343**,** 747-751.
HILL, W. G. 2010. Understanding and using quantitative genetic variation. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 365**,** 73-85.
HIRA, Z. M. & GILLIES, D. F. 2015. A review of feature selection and feature extraction methods applied on microarray data. *Advances in bioinformatics,* 2015.
HOFFMANN, A. A. & SGRO, C. M. 2011. Climate change and evolutionary adaptation. *Nature,* 470**,** 479-485.
HOLLIDAY, J. A., WANG, T. & AITKEN, S. 2012. Predicting adaptive phenotypes from multilocus genotypes in Sitka spruce (Picea sitchensis) using random forest. *G3: Genes, Genomes, Genetics,* 2**,** 1085-1093.
HORNIK, K., STINCHCOMBE, M. & WHITE, H. 1989. Multilayer feedforward networks are universal approximators. *Neural networks,* 2**,** 359-366.
HOTHORN, T., HORNIK, K., STROBL, C. & ZEILEIS, A. 2010. Party: A laboratory for recursive partytioning.
HOTHORN, T., LEISCH, F., ZEILEIS, A. & HORNIK, K. 2005. The design and analysis of benchmark experiments. *Journal of Computational and Graphical Statistics,* 14**,** 675-699.
HUSSAIN, S., DAS, A., NGUYEN, B. P., MARZUKI, M., LIN, S., KUMAR, A., WRIGHT, G. & SINGHAL, A. 2019. DeLHCA: Deep transfer learning for high-content analysis of the effects of drugs on immune cells. 796-801.
JABLONSKI, N. G. 2004. The evolution of human skin and skin color. *Annu. Rev. Anthropol.,* 33**,** 585-623.
JACCARD, P. 1912. The distribution of the flora in the alpine zone. 1. *New phytologist,* 11**,** 37-50.
JIANG, B., WU, T.-Y., ZHENG, C. & WONG, W. H. 2017. Learning summary statistic for approximate Bayesian computation via deep neural network. *Statistica Sinica***,** 1595-1618.
JINMING, Y., AHMADI, M., JULLIEN, G. A. & MILLER, W. C. 1998. Model validation and determination for neural network activation function modeling. 548-551.
JOMBART, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics,* 24**,** 1403-1405.
JOMBART, T., DEVILLARD, S. & BALLOUX, F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC genetics,* 11**,** 94.

JOOST, S., BONIN, A., BRUFORD, M. W., DESPRÉS, L., CONORD, C., ERHARDT, G. & TABERLET, P. 2007. A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Molecular ecology,* 16**,** 3955-3969.
JOST, L. 2008. GST and its relatives do not measure differentiation. *Molecular ecology,* 17**,** 4015-4026.
JOST, L., DEVRIES, P., WALLA, T., GREENEY, H., CHAO, A. & RICOTTA, C. 2010. Partitioning diversity for conservation analyses. *Diversity and Distributions,* 16**,** 65-76.
JUDD, J. S. 1990. Neural network design and the complexity of learning / J. Stephen Judd. Cambridge, Mass. :: MIT Press.
KALININ, A. A., HIGGINS, G. A., REAMAROON, N., SOROUSHMEHR, S. M. R., ALLYN-FEUER, A., DINOV, I. D., NAJARIAN, K. & ATHEY, B. D. 2018. Deep Learning in Pharmacogenomics: From Gene Regulation to Patient Stratification.
KALKATAWI, M., MAGANA-MORA, A., JANKOVIC, B. & BAJIC, V. B. 2019. DeepGSR: an optimized deep-learning structure for the recognition of genomic signals and regions.
KAPRIO, J. & KOSKENVUO, M. 2002. Genetic and environmental factors in complex diseases: the older Finnish Twin Cohort. *Twin Research and Human Genetics,* 5**,** 358-365.
KELLER, S. R., SOWELL, D. R., NEIMAN, M., WOLFE, L. M. & TAYLOR, D. R. 2009. Adaptation and colonization history affect the evolution of clines in two introduced species. *New Phytologist,* 183**,** 678-690.
KERN, A. D. & SCHRIDER, D. R. 2018. diploS/HIC: an updated approach to classifying selective sweeps. *G3: Genes, Genomes, Genetics,* 8**,** 1959-1970.
KIM, H., GRUENEBERG, A., VAZQUEZ, A. I., HSU, S. & DE LOS CAMPOS, G. 2017. Will big data close the missing heritability gap? *Genetics,* 207**,** 1135-1145.
KIMMEL, G., JORDAN, M. I., HALPERIN, E., SHAMIR, R. & KARP, R. M. 2007. A randomization test for controlling population stratification in whole-genome association studies. *The American Journal of Human Genetics,* 81**,** 895-905.
KLIMA, G. 2016. FCNN4R: Fast Compressed Neural Networks for R. *R package version 0.6,* 2.
KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. Ijcai, 1995. Montreal, Canada, 1137-1145.
KOPP, W., MONTI, R., TAMBURRINI, A., OHLER, U. & AKALIN, A. 2020. Deep learning for genomics using Janggu. *Nature communications,* 11**,** 1-7.
KÖPPEN, M. The curse of dimensionality. 5th Online World Conference on Soft Computing in Industrial Applications (WSC5), 2000. 4-8.
KOUMAKIS, L. 2020. Deep learning models in genomics; are we there yet? *Computational and Structural Biotechnology Journal*.
KRIEGESKORTE, N. 2015. Deep neural networks: a new framework for modeling biological vision and brain information processing. *Annual review of vision science,* 1**,** 417-446.
KUHN, M. 2008. Building predictive models in R using the caret package. *Journal of statistical software,* 28**,** 1-26.
KUHN, M. 2012. The caret package. *R Foundation for Statistical Computing, Vienna, Austria. URL https://cran. r-project. org/package= caret*.
KUHN, M. 2014. Futility analysis in the cross-validation of machine learning models. *arXiv preprint arXiv:1405.6974*.
KUHN, M. 2015. Caret: classification and regression training. *Astrophysics Source Code Library*.
KUHN, M. & JOHNSON, K. 2013. *Applied predictive modeling*, Springer.
LABBÉ, N., LEE, S.-H., CHO, H.-W., JEONG, M. K. & ANDRÉ, N. 2008. Enhanced discrimination and calibration of biomass NIR spectral data using non-linear kernel methods. *Bioresource technology,* 99**,** 8445-8452.
LABONTE, R., POLANYI, M., MUHAJARINE, N., MCINTOSH, T. & WILLIAMS, A. 2005. Beyond the divides: Towards critical population health research. *Critical Public Health,* 15**,** 5-17.

LALOË, D. & GAUTIER, M. 2012. *On the genetic interpretation of between-group PCA on SNP data.* auto-saisine.
LANDE, R. 1996. Statistics and partitioning of species diversity, and similarity among multiple communities. *Oikos***,** 5-13.
LAWSON, D. J., HELLENTHAL, G., MYERS, S. & FALUSH, D. 2012. Inference of population structure using dense haplotype data. *PLoS genetics,* 8.
LAWSON, D. J., VAN DORP, L. & FALUSH, D. 2018. A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nature Communications,* 9**,** 1-11.
LEBERG, P. 2002. Estimating allelic richness: effects of sample size and bottlenecks. *Molecular ecology,* 11**,** 2445-2449.
LECUN, Y., BENGIO, Y. & HINTON, G. 2015. Deep learning. *nature,* 521**,** 436-444.
LEGENDRE, P. & LEGENDRE, L. 2012. Canonical analysis. *Developments in Environmental Modelling.* Elsevier.
LEONENKO, G., LOS, S. & NORTH, P. 2013. Statistical distances and their applications to biophysical parameter estimation: Information measures, M-estimates, and minimum contrast methods. *Remote Sensing,* 5**,** 1355-1388.
LEPPÄLÄ, K., NIELSEN, S. V. & MAILUND, T. 2017. admixturegraph: an R package for admixture graph manipulation and fitting. *Bioinformatics,* 33**,** 1738-1740.
LI, H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics,* 27**,** 2987-2993.
LI, L. & TALWALKAR, A. Random search and reproducibility for neural architecture search. Uncertainty in Artificial Intelligence, 2020. PMLR, 367-377.
LI, S., YANG, Z., LI, H. & SHU, G. 2018. Projection of population structure in China using least squares support vector machine in conjunction with a Leslie matrix model. *Journal of Forecasting,* 37**,** 225-234.
LI, W., CERISE, J. E., YANG, Y. & HAN, H. 2017. Application of t-SNE to human genetic data. *Journal of bioinformatics and computational biology,* 15**,** 1750017.
LIAW, A. & WIENER, M. 2002. Classification and regression by randomForest. *R news,* 2**,** 18-22.
LIU, R. & GILLIES, D. F. 2016. Overfitting in linear feature extraction for classification of high-dimensional image data. *Pattern Recognition,* 53**,** 73-86.
LOTTERHOS, K. E. & WHITLOCK, M. C. 2015. The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular ecology,* 24**,** 1031-1046.
LOUADI, Z., OUBOUNYT, M., TAYARA, H. & CHONG, K. T. 2019. Deep splicing code: Classifying alternative splicing events using deep learning. *Genes,* 10**,** 587.
LUIKART, G., ALLENDORF, F., CORNUET, J. & SHERWIN, W. 1998. Distortion of allele frequency distributions provides a test for recent population bottlenecks. *Journal of Heredity,* 89**,** 238-247.
LUO, D. & LIU, A. 2015. Kernel Fisher discriminant analysis based on a regularized method for multiclassification and application in lithological identification. *Mathematical Problems in Engineering,* 2015.
LUO, P., LI, Y., TIAN, L.-P. & WU, F.-X. 2019. Enhancing the prediction of disease–gene associations with multimodal deep learning. *Bioinformatics,* 35**,** 3735-3742.
LUU, K., BAZIN, E. & BLUM, M. G. 2017. pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Molecular ecology resources,* 17**,** 67-77.
MA, L., JI, Y.-J. & ZHANG, D.-X. 2015. Statistical measures of genetic differentiation of populations: Rationales, history and current states. *Current Zoology,* 61**,** 886-897.
MAATEN, L. V. D. & HINTON, G. 2008. Visualizing data using t-SNE. *Journal of machine learning research,* 9**,** 2579-2605.

MACKAY, D. J. & MAC KAY, D. J. 2003. *Information theory, inference and learning algorithms*, Cambridge university press.
MAJI, P. 2019. Recent advances in multimodal big data analysis for cancer diagnosis. *CSI Transactions on ICT,* 7**,** 227-231.
MANICA, A., PRUGNOLLE, F. & BALLOUX, F. 2005. Geography is a better determinant of human genetic differentiation than ethnicity. *Human genetics,* 118**,** 366-371.
MANOLIO, T. A., COLLINS, F. S., COX, N. J., GOLDSTEIN, D. B., HINDORFF, L. A., HUNTER, D. J., MCCARTHY, M. I., RAMOS, E. M., CARDON, L. R. & CHAKRAVARTI, A. 2009. Finding the missing heritability of complex diseases. *Nature,* 461**,** 747-753.
MATEOS, A., DOPAZO, J., JANSEN, R., TU, Y., GERSTEIN, M. & STOLOVITZKY, G. 2002. Systematic learning of gene functional classes from DNA array expression data by using multilayer perceptrons. *Genome Research,* 12**,** 1703-1715.
MCHUGH, M. L. 2012. Interrater reliability: the kappa statistic. *Biochemia medica: Biochemia medica,* 22**,** 276-282.
MCINNES, L., HEALY, J. & MELVILLE, J. 2018. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.
MCVEAN, G. 2009. A genealogical interpretation of principal components analysis. *PLoS genetics,* 5.
MEIRMANS, P. G. & HEDRICK, P. W. 2011. Assessing population structure: FST and related measures. *Molecular ecology resources,* 11**,** 5-18.
MEMON, N., PATEL, S. B. & PATEL, D. P. Comparative analysis of artificial neural network and XGBoost algorithm for PolSAR image classification. International Conference on Pattern Recognition and Machine Intelligence, 2019. Springer, 452-460.
MERL, D. M. 2006. *Detecting Patterns of Natural Selection Using Bayesian Generalized Linear Models*, Citeseer.
MIAO, L., WANG, X. & WANG, Z. 2004. Relationships between species diversity indices of {\sl Larix chinensis} communities and environmental factors. *Acta Botanica Boreali-occidentalia Sinica,* 24**,** 1888-1894.
MIIKKULAINEN, R., LIANG, J., MEYERSON, E., RAWAL, A., FINK, D., FRANCON, O., RAJU, B., SHAHRZAD, H., NAVRUZYAN, A. & DUFFY, N. 2019. Evolving deep neural networks. *Artificial Intelligence in the Age of Neural Networks and Brain Computing.* Elsevier.
MIN, X., CHEN, N., CHEN, T. & JIANG, R. DeepEnhancer: Predicting enhancers by convolutional neural networks. 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2016. IEEE, 637-644.
MINNOYE, L., TASKIRAN, I. I., MAUDUIT, D., FAZIO, M., VAN AERSCHOT, L., HULSEMANS, G., CHRISTIAENS, V., MAKHZAMI, S., SELTENHAMMER, M. & KARRAS, P. 2019. Cross-species analysis of melanoma enhancer logic using deep learning. *bioRxiv*.
MIRA, J. & SÁNCHEZ-ANDRÉS, J. V. 1999. *Engineering Applications of Bio-Inspired Artificial Neural Networks: International Work-Conference on Artificial and Natural Neural Networks, IWANN'99, Alicante, Spain, June 2-4, 1999, Proceedings*, Springer Science & Business Media.
MITCHELL, M. W. 2011. Bias of the Random Forest out-of-bag (OOB) error for certain input parameters. *Open Journal of Statistics,* 1**,** 205.
MOKRY, F. B., HIGA, R. H., DE ALVARENGA MUDADU, M., DE LIMA, A. O., MEIRELLES, S. L. C., DA SILVA, M. V. G. B., CARDOSO, F. F., DE OLIVEIRA, M. M., URBINATI, I. & NICIURA, S. C. M. 2013. Genome-wide association study for backfat thickness in Canchim beef cattle using Random Forest approach. *BMC genetics,* 14**,** 47.
MONEY, D., GARDNER, K., MIGICOVSKY, Z., SCHWANINGER, H., ZHONG, G.-Y. & MYLES, S. 2015. LinkImpute: fast and accurate genotype imputation for nonmodel organisms. *G3: Genes, Genomes, Genetics,* 5**,** 2383-2390.
MORAN, P. A. 1950. Notes on continuous stochastic phenomena. *Biometrika,* 37**,** 17-23.
MORTON, N. 1977. Isolation by distance in human populations. *Annals of human genetics,* 40**,** 361-365.

MOSTAVI, M., CHIU, Y.-C., HUANG, Y. & CHEN, Y. 2020. Convolutional neural network models for cancer type prediction based on gene expression. *BMC Medical Genomics,* 13**,** 1-13.
NAIR, V. & HINTON, G. E. Rectified linear units improve restricted boltzmann machines. ICML, 2010.
NAKAYAMA, K., HIRANO, A. & IDO, I. 1999. A multilayer neural network with nonlinear inputs and trainable activation functions: structure and simultaneous learning algorithm. 3**,** 1657-1661.
NALLS, M. A., PANKRATZ, N., LILL, C. M., DO, C. B., HERNANDEZ, D. G., SAAD, M., DESTEFANO, A. L., KARA, E., BRAS, J. & SHARMA, M. 2014. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nature genetics,* 46**,** 989-993.
NARUM, S. R. & HESS, J. E. 2011. Comparison of FST outlier tests for SNP loci under selection. *Molecular ecology resources,* 11**,** 184-194.
NEI, M. 1973. Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences,* 70**,** 3321-3323.
NEI, M. 1975. *Molecular population genetics and evolution*, North-Holland Publishing Company.
NEI, M., MARUYAMA, T. & CHAKRABORTY, R. 1975. The bottleneck effect and genetic variability in populations. *Evolution,* 29**,** 1-10.
NELSON, M. R., BRYC, K., KING, K. S., INDAP, A., BOYKO, A. R., NOVEMBRE, J., BRILEY, L. P., MARUYAMA, Y., WATERWORTH, D. M. & WAEBER, G. 2008. The Population Reference Sample, POPRES: a resource for population, disease, and pharmacological genetics research. *The American Journal of Human Genetics,* 83**,** 347-358.
NOVEMBRE, J., JOHNSON, T., BRYC, K., KUTALIK, Z., BOYKO, A. R., AUTON, A., INDAP, A., KING, K. S., BERGMANN, S. & NELSON, M. R. 2008. Genes mirror geography within Europe. *Nature,* 456**,** 98-101.
NOVEMBRE, J. & STEPHENS, M. 2008. Interpreting principal component analyses of spatial population genetic variation. *Nature genetics,* 40**,** 646-649.
NYE, J., ZINGARETTI, L. M. & PÉREZ-ENCISO, M. 2020. Estimating conformational traits in dairy cattle with DeepAPS: a two-step Deep learning Automated Phenotyping and Segmentation approach. *Frontiers in Genetics,* 11.
OKSANEN, J., KINDT, R., LEGENDRE, P., O'HARA, B., STEVENS, M. H. H., OKSANEN, M. J. & SUGGESTS, M. 2007. The vegan package. *Community ecology package,* 10**,** 719.
OLDEN, J. D. & JACKSON, D. A. 2002. Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks. *Ecological modelling,* 154**,** 135-150.
OLDEN, J. D., JOY, M. K. & DEATH, R. G. 2004. An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Ecological modelling,* 178**,** 389-397.
OSBORN, G. 1902. 109.[D. 6. d.] Mnemonic for hyperbolic formulae. *The Mathematical Gazette,* 2**,** 189-189.
PAL, M. 2005. Random forest classifier for remote sensing classification. *International journal of remote sensing,* 26**,** 217-222.
PANCHAL, G., GANATRA, A., KOSTA, Y. & PANCHAL, D. 2011. Behaviour analysis of multilayer perceptrons with multiple hidden neurons and hidden layers. *International Journal of Computer Theory and Engineering,* 3**,** 332-337.
PARK, Y. & KELLIS, M. 2015. Deep learning for regulatory genomics. *Nature biotechnology,* 33**,** 825.
PARSONS, T. 1966. *Societies: Evolutionary and comparative perspectives*, Prentice-Hall Englewood Cliffs, NJ.
PARTALAS, I., TSOUMAKAS, G. & VLAHAVAS, I. P. Focused Ensemble Selection: A Diversity-Based Method for Greedy Ensemble Selection. ECAI, 2008. 117-121.
PATTERSON, N., PRICE, A. L. & REICH, D. 2006. Population structure and eigenanalysis. *PLoS genet,* 2**,** e190.

PELTONEN, L., PALOTIE, A. & LANGE, K. 2000. Use of population isolates for mapping complex traits. *Nature Reviews Genetics,* 1**,** 182-190.
PEREZ-ENCISO, M. & ZINGARETTI, L. M. 2019. A Guide on Deep Learning for Complex Trait Genomic Prediction.
PÉREZ-ENCISO, M. & ZINGARETTI, L. M. 2019. A guide on deep learning for complex trait genomic prediction. *Genes,* 10**,** 553.
PEREZ, M. F., FRANCO, F. F., BOMBONATO, J. R., BONATELLI, I. A., KHAN, G., ROMEIRO‐BRITO, M., FEGIES, A. C., RIBEIRO, P. M., SILVA, G. A. & MORAES, E. M. 2018. Assessing population structure in the face of isolation by distance: Are we neglecting the problem? *Diversity and Distributions,* 24**,** 1883-1889.
PETER, B. M., PETKOVA, D. & NOVEMBRE, J. 2020. Genetic landscapes reveal how human genetic diversity aligns with geography. *Molecular biology and evolution,* 37**,** 943-951.
PICKRELL, J. & PRITCHARD, J. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. *Nature Precedings***,** 1-1.
PINAYA, W. H. L., VIEIRA, S., GARCIA-DIAS, R. & MECHELLI, A. 2020. Convolutional neural networks. *Machine Learning.* Elsevier.
PLATZER, A. 2013. Visualization of SNPs with t-SNE. *PloS one,* 8**,** e56883.
PLOMIN, R. 2014. Genotype-environment correlation in the era of DNA. *Behavior genetics,* 44**,** 629-638.
POLLASTRI, G., PRZYBYLSKI, D., ROST, B. & BALDI, P. 2002. Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles. *Proteins: Structure, Function, and Bioinformatics,* 47**,** 228-235.
PRICE, A. L., PATTERSON, N. J., PLENGE, R. M., WEINBLATT, M. E., SHADICK, N. A. & REICH, D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics,* 38**,** 904-909.
PRITCHARD, J. K. & DI RIENZO, A. 2010. Adaptation–not by sweeps alone. *Nature Reviews Genetics,* 11**,** 665-667.
PRITCHARD, J. K., STEPHENS, M. & DONNELLY, P. 2000. Inference of population structure using multilocus genotype data. *Genetics,* 155**,** 945-959.
PRIVÉ, F., LUU, K., VILHJÁLMSSON, B. J. & BLUM, M. G. 2020. Performing highly efficient genome scans for local adaptation with R package pcadapt version 4. *Molecular Biology and Evolution*.
PRUGNOLLE, F., MANICA, A. & BALLOUX, F. 2005. Geography predicts neutral genetic diversity of human populations. *Current Biology,* 15**,** R159-R160.
PUECHMAILLE, S. J. 2016. The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. *Molecular ecology resources,* 16**,** 608-627.
QIAN, Y., FAN, Y., HU, W. & SOONG, F. K. On the training aspects of deep neural network (DNN) for parametric TTS synthesis. 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2014. IEEE, 3829-3833.
QIN, X. 2020. KLFDAPC v0. 2.0.
RABUÑAL, J. R. 2005. *Artificial neural networks in real-life applications*, IGI Global.
RAJ, A., STEPHENS, M. & PRITCHARD, J. K. 2014. fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics,* 197**,** 573-589.
RAMEZANKHANI, A., POURNIK, O., SHAHRABI, J., KHALILI, D., AZIZI, F. & HADAEGH, F. 2014. Applying decision tree for identification of a low risk population for type 2 diabetes. Tehran Lipid and Glucose Study. *Diabetes research and clinical practice,* 105**,** 391-398.
RAY, B., HENAFF, M., MA, S., EFSTATHIADIS, E., PESKIN, E. R., PICONE, M., POLI, T., ALIFERIS, C. F. & STATNIKOV, A. 2014. Information content and analysis methods for multi-modal high-throughput biomedical data. *Scientific reports,* 4**,** 4411.

REED, R. & MARKS, I. 1989. RJ (1999). Neural Smithing. Supervised Learning in Feedforward Artificial Neural Networks. *Cambridge, MA: MIT Press*.
REICH, D., PRICE, A. L. & PATTERSON, N. 2008. Principal component analysis of genetic data. *Nature genetics,* 40**,** 491-492.
RELLSTAB, C., GUGERLI, F., ECKERT, A. J., HANCOCK, A. M. & HOLDEREGGER, R. 2015. A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology,* 24**,** 4348-4370.
RIBEIRO, M. T., SINGH, S. & GUESTRIN, C. 2016. Model-agnostic interpretability of machine learning. *arXiv preprint arXiv:1606.05386*.
RIBEIRO, M. T., SINGH, S. & GUESTRIN, C. Anchors: High-precision model-agnostic explanations. Thirty-Second AAAI Conference on Artificial Intelligence, 2018.
RICOTTA, C. 2005. On hierarchical diversity decomposition. *Journal of Vegetation Science,* 16**,** 223-226.
RIEBLER, A., HELD, L. & STEPHAN, W. 2008. Bayesian variable selection for detecting adaptive genomic differences among populations. *Genetics,* 178**,** 1817-1829.
RIEDMILLER, M. 1994. Advanced supervised learning in multi-layer perceptrons—from backpropagation to adaptive learning algorithms. *Computer Standards & Interfaces,* 16**,** 265-278.
RIEDMILLER, M. & BRAUN, H. A direct adaptive method for faster backpropagation learning: The RPROP algorithm. IEEE international conference on neural networks, 1993. IEEE, 586-591.
RIESSELMAN, A. J., INGRAHAM, J. B. & MARKS, D. S. 2017. Deep generative models of genetic variation capture mutation effects.
RIPLEY, B., VENABLES, B., BATES, D. M., HORNIK, K., GEBHARDT, A., FIRTH, D. & RIPLEY, M. B. 2013. Package 'mass'. *Cran R,* 538.
RIPLEY, B., VENABLES, W. & RIPLEY, M. B. 2016. Package 'nnet'. *R package version,* 7**,** 3-12.
RIPLEY, B. D. & HJORT, N. 1996. *Pattern recognition and neural networks*, Cambridge university press.
RIVENSON, Y., ZHANG, Y., GÜNAYDıN, H., TENG, D. & OZCAN, A. 2018. Phase recovery and holographic image reconstruction using deep learning in neural networks. *Light: Science & Applications,* 7**,** 17141.
ROBERTS, A., MCMILLAN, L., WANG, W., PARKER, J., RUSYN, I. & THREADGILL, D. 2007. Inferring missing genotypes in large SNP panels using fast nearest-neighbor searches over sliding windows. *Bioinformatics,* 23**,** i401-i407.
ROBINSON, A. H. Which map is best?: projections for world maps. 1986. Amer Congress on Surveying &.
ROCKMAN, M. V. 2012. The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution: International Journal of Organic Evolution,* 66**,** 1-17.
ROOT, M. 2000. How we divide the world. *Philosophy of Science,* 67**,** S628-S639.
ROSENBERG, N. A., MAHAJAN, S., RAMACHANDRAN, S., ZHAO, C., PRITCHARD, J. K. & FELDMAN, M. W. 2005. Clines, clusters, and the effect of study design on the inference of human population structure. *PLoS Genet,* 1**,** e70.
ROULLIER, C., BENOIT, L., MCKEY, D. B. & LEBOT, V. 2013. Historical collections reveal patterns of diffusion of sweet potato in Oceania obscured by modern plant movements and recombination. *Proceedings of the National Academy of Sciences,* 110**,** 2205-2210.
RUFF, C. 2002. Variation in human body size and shape. *Annual Review of Anthropology,* 31**,** 211-232.
RUMELHART, D. E., HINTON, G. E. & WILLIAMS, R. J. 1986. Learning representations by back-propagating errors. *nature,* 323**,** 533-536.
SABETI, P. C., REICH, D. E., HIGGINS, J. M., LEVINE, H. Z., RICHTER, D. J., SCHAFFNER, S. F., GABRIEL, S. B., PLATKO, J. V., PATTERSON, N. J. & MCDONALD, G. J. 2002. Detecting recent positive selection in the human genome from haplotype structure. *Nature,* 419**,** 832-837.

SAEB, A. & AL-NAQEB, D. 2016. The impact of evolutionary driving forces on human complex diseases: a population genetics approach. *Scientifica,* 2016.
SAETHER, S. A., FISKE, P., KÅLÅS, J., KURESOO, A., LUIGUJOE, L., PIERTNEY, S., SAHLMAN, T. & HÖGLUND, J. 2007. Inferring local adaptation from QST–FST comparisons: neutral genetic and quantitative trait variation in European populations of great snipe. *Journal of Evolutionary Biology,* 20**,** 1563-1576.
SALEHINEJAD, H., SANKAR, S., BARFETT, J., COLAK, E. & VALAEE, S. 2017. Recent advances in recurrent neural networks. *arXiv preprint arXiv:1801.01078*.
SAMWORTH, R. J. 2014. Big Data: a new era for Statistics. *The Eagle, Bryan***,** 43-46.
SCHAFFER, C. 1993. Selecting a classification method by cross-validation. *Machine Learning,* 13**,** 135-143.
SCHMIDHUBER, J. 2015. Deep learning in neural networks: An overview. *Neural networks,* 61**,** 85-117.
SCHOEN, D. J. & BROWN, A. 1993. Conservation of allelic richness in wild crop relatives is aided by assessment of genetic markers. *Proceedings of the national academy of sciences,* 90**,** 10623-10627.
SCHRIDER, D. R. & KERN, A. D. 2016a. S/HIC: robust identification of soft and hard sweeps using machine learning. *PLoS genetics,* 12**,** e1005928.
SCHRIDER, D. R. & KERN, A. D. 2016b. S/HIC: Robust Identification of Soft and Hard Sweeps Using Machine Learning. 12**,** 1-31.
SCHRIDER, D. R. & KERN, A. D. 2018. Supervised machine learning for population genetics: a new paradigm. *Trends in Genetics,* 34**,** 301-312.
SCHWENDER, H. 2012. Imputing missing genotypes with weighted k nearest neighbors. *Journal of Toxicology and Environmental Health, Part A,* 75**,** 438-446.
SCOTT, J. M., CSUTI, B., JACOBI, J. D. & ESTES, J. E. 1987. Species richness. *BioScience,* 37**,** 782-788.
SEGAL, M. R. 2004. Machine learning benchmarks and random forest regression.
SEJNOWSKI, T. J. 2020. The Unreasonable Effectiveness of Deep Learning in Artificial Intelligence.
SEKHON, A., SINGH, R. & QI, Y. 2018. DeepDiff: Deep-learning for predicting Differential gene expression from histone modifications.
SERRE, D. & PÄÄBO, S. 2004. Evidence for gradients of human genetic diversity within and among continents. *Genome research,* 14**,** 1679-1685.
SHARMA, C., SUYAL, S., GAIROLA, S. & GHILDIYAL, S. 2009. Species richness and diversity along an altitudinal gradient in moist temperate forest of Garhwal Himalaya. *Journal of American Science,* 5**,** 119-128.
SHARMA, S. 2017. Activation functions in neural networks. *Towards Data Science,* 6.
SHEEHAN, S. & SONG, Y. S. 2016. Deep learning for population genetic inference. *PLoS computational biology,* 12**,** e1004845.
SHEN, H., WELCH, W. J. & HUGHES-OLIVER, J. M. 2011. Efficient, adaptive cross-validation for tuning and comparing models, with application to drug discovery. *The Annals of applied statistics***,** 2668-2687.
SHERWIN, W. B. 2018. Entropy, or Information, Unifies Ecology and Evolution and Beyond. *Entropy,* 20**,** 727.
SHERWIN, W. B., CHAO, A., JOST, L. & SMOUSE, P. E. 2017. Information theory broadens the spectrum of molecular ecology and evolution. *Trends in ecology & evolution,* 32**,** 948-963.
SHRIVER, M. D., MEI, R., PARRA, E. J., SONPAR, V., HALDER, I., TISHKOFF, S. A., SCHURR, T. G., ZHADANOV, S. I., OSIPOVA, L. P. & BRUTSAERT, T. D. 2005. Large-scale SNP analysis reveals clustered and continuous patterns of human genetic variation. *Human genomics,* 2**,** 1-9.
SILLIMAN, K. 2019. Population structure, genetic connectivity, and adaptation in the Olympia oyster (Ostrea lurida) along the west coast of North America. *Evolutionary applications,* 12**,** 923-939.

SINGH, R., LANCHANTIN, J., ROBINS, G. & QI, Y. 2016. DeepChrome: deep-learning for predicting gene expression from histone modifications. *Bioinformatics,* 32**,** i639-i648.
SLOAN, C. D., DUELL, E. J., SHI, X., IRWIN, R., ANDREW, A. S., WILLIAMS, S. M. & MOORE, J. H. 2009. Ecogeographic genetic epidemiology. *Genetic Epidemiology: The Official Publication of the International Genetic Epidemiology Society,* 33**,** 281-289.
SNAPINN, S., CHEN, M. G., JIANG, Q. & KOUTSOUKOS, T. 2006. Assessment of futility in clinical trials. *Pharmaceutical Statistics: The Journal of Applied Statistics in the Pharmaceutical Industry,* 5**,** 273-281.
STOTZ, G. C., GIANOLI, E. & CAHILL, J. F. 2016. Spatial pattern of invasion and the evolutionary responses of native plant species. *Evolutionary Applications,* 9**,** 939-951.
STROBL, C., BOULESTEIX, A.-L., KNEIB, T., AUGUSTIN, T. & ZEILEIS, A. 2008. Conditional variable importance for random forests. *BMC bioinformatics,* 9**,** 307.
STROBL, C., BOULESTEIX, A.-L., ZEILEIS, A. & HOTHORN, T. 2007. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC bioinformatics,* 8**,** 25.
SU, Y., WEI, W., ROBERT, L., XUE, M., TSOI, J., GARCIA-DIAZ, A., MORENO, B. H., KIM, J., NG, R. H. & LEE, J. W. 2017. Single-cell analysis resolves the cell state transition and signaling dynamics associated with melanoma drug-induced resistance. *Proceedings of the National Academy of Sciences,* 114**,** 13679-13684.
SUGIYAMA, M. Local fisher discriminant analysis for supervised dimensionality reduction. Proceedings of the 23rd international conference on Machine learning, 2006. 905-912.
SUGIYAMA, M. 2007. Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. *Journal of machine learning research,* 8**,** 1027-1061.
SVETNIK, V., LIAW, A., TONG, C., CULBERSON, J. C., SHERIDAN, R. P. & FEUSTON, B. P. 2003. Random forest: a classification and regression tool for compound classification and QSAR modeling. *Journal of chemical information and computer sciences,* 43**,** 1947-1958.
SYLVESTER, E. V., BENTZEN, P., BRADBURY, I. R., CLÉMENT, M., PEARCE, J., HORNE, J. & BEIKO, R. G. 2018. Applications of random forest feature selection for fine-scale genetic population assignment. *Evolutionary applications,* 11**,** 153-165.
TABERLET, P., ZIMMERMANN, N. E., ENGLISCH, T., TRIBSCH, A., HOLDEREGGER, R., ALVAREZ, N., NIKLFELD, H., COLDEA, G., MIREK, Z. & MOILANEN, A. 2012. Genetic diversity in widespread species is not congruent with species richness in alpine plant communities. *Ecology Letters,* 15**,** 1439-1448.
TANG, Y. & LI, W. 2016. Lfda: An R package for local fisher discriminant analysis and visualization. *arXiv preprint arXiv:1612.09219*.
TANG, Y. & LI, W. 2017. lfda: Local Fisher Discriminant Analysis in R.
TEAM, R. C. 2013. R: A language and environment for statistical computing. Vienna, Austria.
THOMPSON, N. 2016. *Anti-discriminatory practice: Equality, diversity and social justice*, Macmillan International Higher Education.
TORADA, L., LORENZON, L., BEDDIS, A., ISILDAK, U., PATTINI, L., MATHIESON, S. & FUMAGALLI, M. 2019. ImaGene: a convolutional neural network to quantify natural selection from genomic data.
TSUMURA, Y., KADO, T., TAKAHASHI, T., TANI, N., UJINO-IHARA, T. & IWATA, H. 2007. Genome scan to detect genetic structure and adaptive genes of natural populations of Cryptomeria japonica. *Genetics,* 176**,** 2393-2403.
ÜSTÜN, B., MELSSEN, W., OUDENHUIJZEN, M. & BUYDENS, L. 2005. Determination of optimal support vector regression parameters by genetic algorithms and simplex optimization. *Analytica Chimica Acta,* 544**,** 292-305.
VAN HILTEN, A., KUSHNER, S. A., KAYSER, M., IKRAM, M. A., ADAMS, H., KLAVER, C., NIESSEN, W. J. & ROSHCHUPKIN, G. V. 2020. GenNet framework: interpretable neural networks for phenotype prediction. *bioRxiv*.

VAN TIENDEREN, P. H. 1991. Evolution of generalists and specialist in spatially heterogeneous environments. *Evolution*, 1317-1331.
VAPNIK, V. 1998. The support vector method of function estimation. *Nonlinear Modeling.* Springer.
VASSEUR, E. & QUINTANA‐MURCI, L. 2013. The impact of natural selection on health and disease: uses of the population genetics approach in humans. *Evolutionary applications,* 6, 596-607.
VELLEND, M. 2016. *The theory of ecological communities (MPB-57)*, Princeton University Press.
VERITY, R. & NICHOLS, R. A. 2014. What is genetic differentiation, and how should we measure it—GST, D, neither or both? *Molecular ecology,* 23, 4216-4225.
VERT, J.-P., TSUDA, K. & SCHÖLKOPF, B. 2004. A primer on kernel methods. *Kernel methods in computational biology,* 47, 35-70.
VITALIS, R., DAWSON, K. & BOURSOT, P. 2001. Interpretation of variation across marker loci as evidence of selection. *Genetics,* 158, 1811-1823.
VONESH, E. & CHINCHILLI, V. 1997. Linear and Non-Linear Models for the Analysis of Repeated Measurements Marcel Decker. *Inc, New York, NY*.
WAGNER, A. 2017. Information theory, evolutionary innovations and evolvability. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 372, 20160416.
WAGNER, J. E. 2000. Regional economic diversity: action, concept, or state of confusion. *Journal of Regional analysis and policy,* 30.
WAGNER, J. E. & DELLER, S. C. 1993. A measure of economic diversity: An input-output approach. *Staff Paper,* 93.
WANG, C., SZPIECH, Z. A., DEGNAN, J. H., JAKOBSSON, M., PEMBERTON, T. J., HARDY, J. A., SINGLETON, A. B. & ROSENBERG, N. A. 2010. Comparing spatial maps of human population-genetic variation using Procrustes analysis. *Statistical applications in genetics and molecular biology,* 9.
WANG, J., HUANG, J. & YAU, S. S. 2000. Approximate nonlinear output regulation based on the universal approximation theorem. *International Journal of Robust and Nonlinear Control: IFAC‐Affiliated Journal,* 10, 439-456.
WANG, R., FAN, Q., ZHANG, J., ZHANG, X., KANG, Y. & WANG, Z. 2018. Hydrogen sulfide demonstrates promising antitumor efficacy in gastric carcinoma by targeting MGAT5. *Translational oncology,* 11, 900-910.
WANG, X., WIEGAND, T., WOLF, A., HOWE, R., DAVIES, S. J. & HAO, Z. 2011. Spatial patterns of tree species richness in two temperate forests. *Journal of Ecology,* 99, 1382-1393.
WANG, X., XUAN, H., EVERS, B., SHRESTHA, S., PLESS, R. & POLAND, J. 2019. High-throughput phenotyping with deep learning gives insight into the genetic architecture of flowering time in wheat.
WANG, Z. & SUN, X. 2013. Multiple kernel local Fisher discriminant analysis for face recognition. *Signal processing,* 93, 1496-1509.
WAPLES, R. S. & GAGGIOTTI, O. 2006. INVITED REVIEW: What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular ecology,* 15, 1419-1439.
WAPLES, R. S., PUNT, A. E. & COPE, J. M. 2008. Integrating genetic data into management of marine resources: how can we do it better? *Fish and Fisheries,* 9, 423-449.
WATTENBERG, M., VIÉGAS, F. & JOHNSON, I. 2016. How to use t-SNE effectively. *Distill,* 1, e2.
WEGMANN, D., LEUENBERGER, C. & EXCOFFIER, L. 2009. Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. *Genetics,* 182, 1207-1218.
WEINBERGER, K. Q., BLITZER, J. & SAUL, L. K. Distance metric learning for large margin nearest neighbor classification. Advances in neural information processing systems, 2006. 1473-1480.
WESTON, J., SCHÖLKOPF, B., ESKIN, E., LESLIE, C. & NOBLE, W. S. 2003. Dealing with large diagonals in kernel matrices. *Annals of the Institute of Statistical Mathematics,* 55, 391-408.

WETTSCHERECK, D. 1994. A study of distance-based machine learning algorithms.
WHEELER, D. L., CHURCH, D. M., FEDERHEN, S., LASH, A. E., MADDEN, T. L., PONTIUS, J. U., SCHULER, G. D., SCHRIML, L. M., SEQUEIRA, E. & TATUSOVA, T. A. 2003. Database resources of the National Center for Biotechnology. *Nucleic acids research,* 31**,** 28-33.
WHITLOCK, M. C., LOTTERHOS, K. E. & BRONSTEIN, J. L. 2015. Reliable Detection of Loci Responsible for Local Adaptation: Inference of a Null Model through Trimming the Distribution of FST. 186**,** S24.
WIKIPEDIA 2020. List of countries and dependencies by area. *In:* WIKIPEDIA (ed.) *The Free Encyclopedia.*
WILKINSON, S., HALEY, C., ALDERSON, L. & WIENER, P. 2011. An empirical assessment of individual-based population genetic statistical techniques: application to British pig breeds. *Heredity,* 106**,** 261-269.
WILSON, J. F., WEALE, M. E., SMITH, A. C., GRATRIX, F., FLETCHER, B., THOMAS, M. G., BRADMAN, N. & GOLDSTEIN, D. B. 2001. Population genetic structure of variable drug response. *Nature genetics,* 29**,** 265-269.
WINKLER, D. A. & LE, T. C. 2017. Performance of deep and shallow neural networks, the universal approximation theorem, activity cliffs, and QSAR. *Molecular informatics,* 36**,** 1600118.
WITOLD, P. & SHYI-MING, C. 2020a. Deep learning : algorithms and applications. Cham, Switzerland :: Springer.
WITOLD, P. & SHYI-MING, C. 2020b. Deep Learning : Concepts and Architectures. Cham, Switzerland :: Springer.
WOLD, S., ESBENSEN, K. & GELADI, P. 1987. Principal component analysis. *Chemometrics and intelligent laboratory systems,* 2**,** 37-52.
WRIGHT, M. N., ZIEGLER, A. & KÖNIG, I. R. 2016. Do little interactions get lost in dark random forests? **,** 1-10.
WRIGHT, S. 1949. The genetical structure of populations. *Annals of eugenics,* 15**,** 323-354.
XIANG, L., STILLWELL, J., BURNS, L. & HEPPENSTALL, A. 2020. Measuring and Assessing Regional Education Inequalities in China under Changing Policy Regimes. *Applied Spatial Analysis and Policy,* 13**,** 91-112.
XUE, A. T., SCHRIDER, D. R., KERN, A. D. & CONSORTIUM, A. 2019. Discovery of ongoing selective sweeps within Anopheles mosquito populations using deep learning. *bioRxiv***,** 589069.
YANG, J., MA, A., HOPPE, A. D., WANG, C., LI, Y., ZHANG, C., WANG, Y., LIU, B. & MA, Q. 2019. Prediction of regulatory motifs from human Chip-sequencing data using a deep learning framework.
YANG, L. 2011. Classifiers selection for ensemble learning based on accuracy and diversity. *Procedia Engineering,* 15**,** 4266-4270.
YANG, W.-Y., NOVEMBRE, J., ESKIN, E. & HALPERIN, E. 2012. A model-based approach for analysis of spatial structure in genetic data. *Nature genetics,* 44**,** 725.
YANG, W.-Y., PLATT, A., CHIANG, C. W.-K., ESKIN, E., NOVEMBRE, J. & PASANIUC, B. 2014a. Spatial localization of recent ancestors for admixed individuals. *G3: Genes, Genomes, Genetics,* 4**,** 2505-2518.
YANG, Z., PELTONEN, J. & KASKI, S. Optimization equivalence of divergences improves neighbor embedding. International Conference on Machine Learning, 2014b. 460-468.
ZAR, J. H. 2005. Spearman rank correlation. *Encyclopedia of biostatistics,* 7.
ZELNIK-MANOR, L. & PERONA, P. 2004. Self-tuning spectral clustering. *Advances in neural information processing systems,* 17**,** 1601-1608.
ZHANG, G. & JIA, X. Feature selection using kernel based local fisher discriminant analysis for hyperspectral image classification. 2011 IEEE International Geoscience and Remote Sensing Symposium, 2011. IEEE, 1728-1731.
ZHANG, L., CHEN, J. X. & LI, S. C. 2019. A multimodal framework for detecting direct and indirect gene-gene interactions from large expression compendium. *bioRxiv***,** 680116.

ZHAO, Y., CHEN, F., ZHAI, R., LIN, X., WANG, Z., SU, L. & CHRISTIANI, D. C. 2012. Correction for population stratification in random forest analysis. *International journal of epidemiology,* 41**,** 1798-1806.

ZHENG, H., YANG, Z., LIU, W., LIANG, J. & LI, Y. Improving deep neural networks using softplus units. 2015 International Joint Conference on Neural Networks (IJCNN), 2015. IEEE, 1-4.

ZHENG, X., LEVINE, D., SHEN, J., GOGARTEN, S. M., LAURIE, C. & WEIR, B. S. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics,* 28**,** 3326-3328.

ZHENG, X. & ZHENG, M. X. 2013. Package 'SNPRelate'.

ZHONGDA, L., JIANHAO, L. & MEIJIN, W. 2015. High-Dimensional Statistics in Big Data Era: Development and Application of Sparse Modeling. *Statistical Research***,** 1.

ZINGARETTI, L. M., GEZAN, S. A., FERRÃO, L. F. V., OSORIO, L. F., MONFORT, A., MUÑOZ, P. R., WHITAKER, V. M. & PÉREZ-ENCISO, M. 2020. Exploring deep learning for complex trait genomic prediction in polyploid outcrossing species. *Frontiers in plant science,* 11**,** 25.

ZOU, J., HUSS, M., ABID, A., MOHAMMADI, P., TORKAMANI, A. & TELENTI, A. 2019. A primer on deep learning in genomics. *Nature genetics,* 51**,** 12-18.